

Estarización de variables

Paúl Arévalo, Esteban Vizhñay

2024-06-28

Estandarización de variables aleatorias

Definición del conjunto de datos

Tenemos una muestra de las alturas de un grupo de personas

```
alturas <- c(
  1.64, 1.60, 1.54, 1.55,
  1.73, 1.61, 1.65, 1.73,
  1.55, 1.61, 1.62, 1.73,
  1.56, 1.65, 1.66, 1.75,
  1.57, 1.63, 1.76, 1.67,
  1.58, 1.79, 1.70, 1.78,
  1.58, 1.63, 1.69, 1.63,
  1.52, 1.59, 1.62, 1.71,
  1.64, 1.53, 1.60, 1.72,
  1.54, 1.60, 1.64, 1.72
)
alturas

## [1] 1.64 1.60 1.54 1.55 1.73 1.61 1.65 1.73 1.55 1.61 1.62 1.73 1.56 1.65 1.66
## [16] 1.75 1.57 1.63 1.76 1.67 1.58 1.79 1.70 1.78 1.58 1.63 1.69 1.63 1.52 1.59
## [31] 1.62 1.71 1.64 1.53 1.60 1.72 1.54 1.60 1.64 1.72
```

Frecuencias

```
frecuencias <- table(alturas)
frecuencias

## alturas
## 1.52 1.53 1.54 1.55 1.56 1.57 1.58 1.59 1.6 1.61 1.62 1.63 1.64 1.65 1.66 1.67
## 1 1 2 2 1 1 2 1 3 2 2 3 3 2 1 1
## 1.69 1.7 1.71 1.72 1.73 1.75 1.76 1.78 1.79
## 1 1 1 2 3 1 1 1 1
```

Media de las estaturas

```
media <- mean(alturas)
media
```

```
## [1] 1.6405
```

Varianza de las estaturas

```
desviacion_estandar_alturas <- sd(alturas)
desviacion_estandar_alturas
```

```
## [1] 0.07330827
```

Tipificación de las alturas

```
tipificacion_alturas <- c((alturas - media) / desviacion_estandar_alturas)
tipificacion_alturas
```

```
## [1] -0.006820513 -0.552461538 -1.370923075 -1.234512819 1.220871793
## [6] -0.416051281 0.129589743 1.220871793 -1.234512819 -0.416051281
## [11] -0.279641025 1.220871793 -1.098102563 0.129589743 0.266000000
## [16] 1.493692306 -0.961692306 -0.143230769 1.630102562 0.402410256
## [21] -0.825282050 2.039333331 0.811641025 1.902923074 -0.825282050
## [26] -0.143230769 0.675230768 -0.143230769 -1.643743588 -0.688871794
## [31] -0.279641025 0.948051281 -0.006820513 -1.507333331 -0.552461538
## [36] 1.084461537 -1.370923075 -0.552461538 -0.006820513 1.084461537
```

Obtención de la media y la desviación estándar

```
sprintf("%.7f", mean(tipificacion_alturas))
```

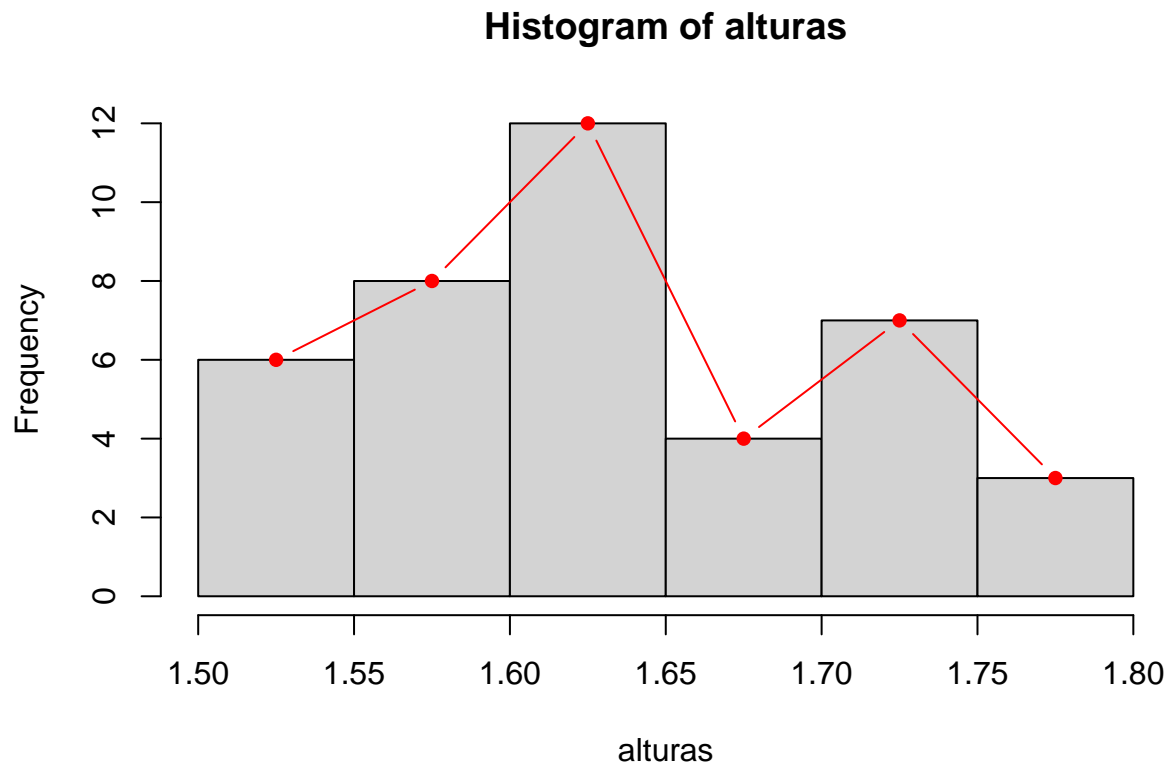
```
## [1] "-0.0000000"
```

```
sd(tipificacion_alturas)
```

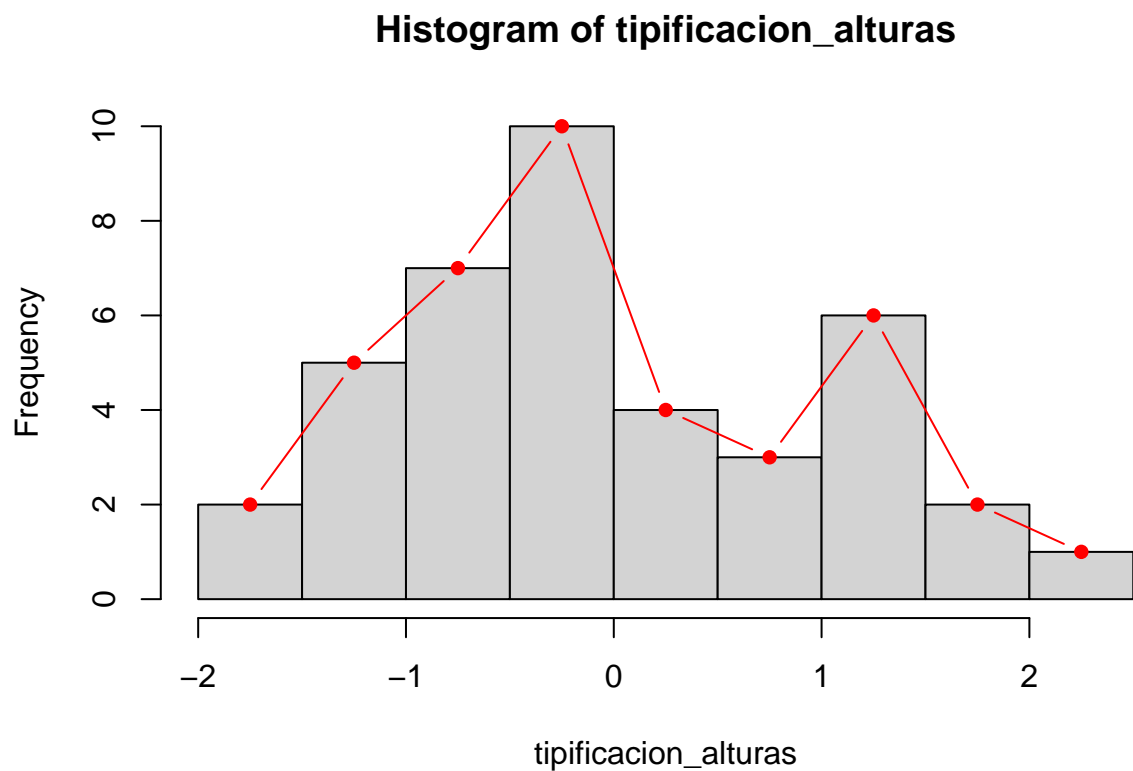
```
## [1] 1
```

Histogramas

```
h <- hist(alturas)
# Calcular los centros de los rectángulos del histograma
centers <- h$mids
# Agregar la línea a través de los centros de los rectángulos
lines(centers, h$counts, type = "b", col = "red", pch = 16)
```



```
h_t <- hist(tipificacion_alturas)
# Calcular los centros de los rectángulos del histograma
centers_t <- h_t$mids
# Agregar la línea a través de los centros de los rectángulos
lines(centers_t, h_t$counts, type = "b", col = "red", pch = 16)
```



Estandarización de las alturas

```
estandarizaciones <- scale(alturas)
estandarizaciones
```

```
##           [,1]
## [1,] -0.006820513
## [2,] -0.552461538
## [3,] -1.370923075
## [4,] -1.234512819
## [5,]  1.220871793
## [6,] -0.416051281
## [7,]  0.129589743
## [8,]  1.220871793
## [9,] -1.234512819
## [10,] -0.416051281
## [11,] -0.279641025
## [12,]  1.220871793
## [13,] -1.098102563
## [14,]  0.129589743
## [15,]  0.266000000
## [16,]  1.493692306
## [17,] -0.961692306
## [18,] -0.143230769
## [19,]  1.630102562
## [20,]  0.402410256
## [21,] -0.825282050
## [22,]  2.039333331
## [23,]  0.811641025
## [24,]  1.902923074
## [25,] -0.825282050
## [26,] -0.143230769
## [27,]  0.675230768
## [28,] -0.143230769
## [29,] -1.643743588
## [30,] -0.688871794
## [31,] -0.279641025
## [32,]  0.948051281
## [33,] -0.006820513
## [34,] -1.507333331
## [35,] -0.552461538
## [36,]  1.084461537
## [37,] -1.370923075
## [38,] -0.552461538
## [39,] -0.006820513
## [40,]  1.084461537
## attr(,"scaled:center")
## [1] 1.6405
## attr(,"scaled:scale")
## [1] 0.07330827
```

Gráfico de histograma para las alturas

```
hist(alturas,  
     col = "blue", border = "black", breaks = 5,  
     main = "Histograma de la distribución de las alturas",  
     xlab = "Alturas", ylab = "Frecuencias",  
     )
```

Histograma de la distribución de las alturas

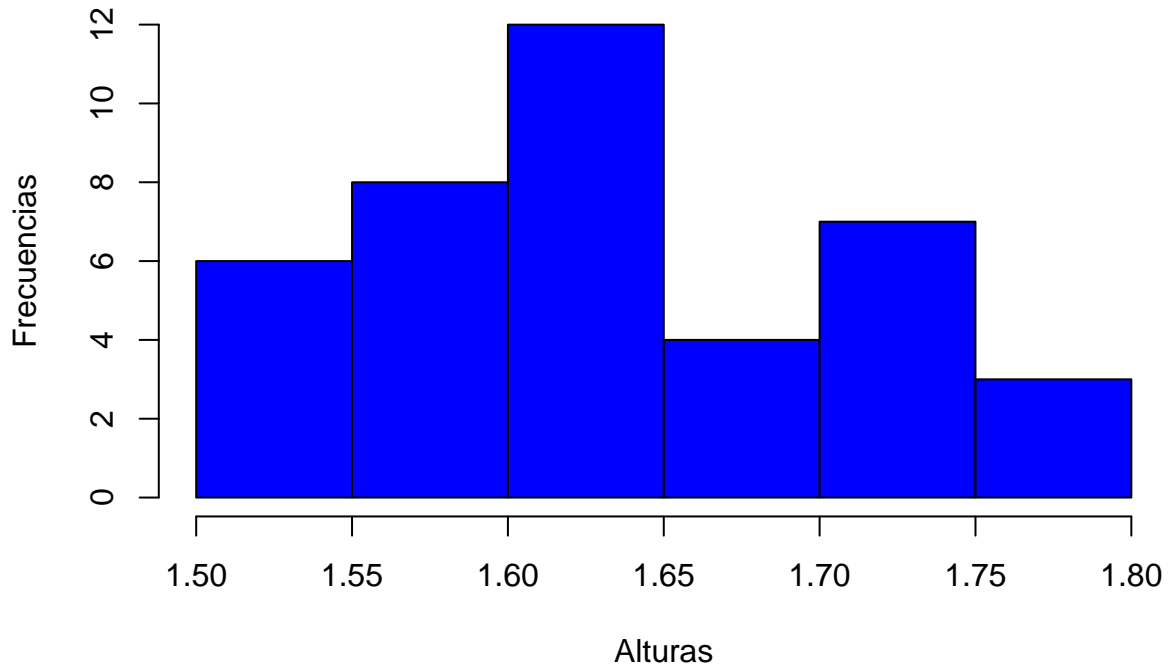


Gráfico de histograma para las alturas normalizadas

```
hist(estandarizaciones,  
     col = "blue", border = "black",  
     main = "Histograma de la distribución de las alturas",  
     xlab = "Alturas", ylab = "Frecuencias",  
     )
```

Histograma de la distribución de las alturas

